UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,174	10/31/2000	Charu C. Aggarwal	YOR920000430US1	7445

7590   11/08/2007

William E Lewis
Ryan Mason & Lewis LLP
90 Forest Avenue
Locust Valley, NY 11560

EXAMINER
HILLERY, NATHAN

ART UNIT	PAPER NUMBER
2176	

MAIL DATE	DELIVERY MODE
11/08/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

PTOL-90A (Rev. 04/07)

	Application No.	Applicant(s)
Office Action Summary	09/703,174	AGGARWAL ET AL.
	Examiner	**Art Unit**	
	Nathan Hillery	2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on <u>27 August 2007</u>.

2a) ☒ This action is **FINAL.**   2b) ☐ This action is non-final.

3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) <u>1-27</u> is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) ☐ Claim(s) _____ is/are allowed.

6) ☒ Claim(s) <u>1-27</u> is/are rejected.

7) ☐ Claim(s) _____ is/are objected to.

8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☐ The specification is objected to by the Examiner.

10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☐ All   b) ☐ Some * c) ☐ None of:

      1. ☐ Certified copies of the priority documents have been received.

      2. ☐ Certified copies of the priority documents have been received in Application No. _____.

      3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

1.   This action is responsive to communications: Amendment filed on 8/27/07.

2.   Claims 1 – 27 are pending in the case. Claims 1, 10, and 19 are independent.

Claim Rejections - 35 USC § 103

3.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4.   Claims 1 – 8, 10 – 17 and 19 – 26 are rejected under 35 U.S.C. 103(a) as being

unpatentable over Chakrabarti et al. (Focused Crawling: A New Approach to Topic-

specific Web Resource Discovery) [as cited by applicant] and in further view of

Chaudhuri et al. (US 6529901 B1).

5.   **Regarding independent claim 1**, Chakrabarti et al. teach that *keyword search is*

used to locate an initial set of pages (using a giant crawl and index) (p 6, section 2.2,

last paragraph), which meet the limitation of **initially retrieving one or more**

documents from the information network that satisfy a user-defined predicate,

wherein the initial document retrieval operation is performed without assuming a

specific model of a linkage structure such that the initial document retrieval

operation retrieves the one or more documents without assuming that a

relationship exists between a feature of a first one of the one or more documents

and a feature of at least another one of the one or more documents that links to

the first one.

6.   Chakrabarti et al. teach that while fetching a document, the above formulation is used to find the leaf node with the highest probability. If some ancestor has been marked good we allow future visitation of URLs found on the document, otherwise the crawl is pruned there (p 9, section Hard focus rule), which meet the limitation of **collecting statistical information about the one or more retrieved documents as the one or more retrieved documents are analyzed and using the collected statistical information to automatically determine further document retrieval operations to be performed in accordance with the information network**, since the probabilities are calculated to find the "best" leaf node, the ancestors are analyzed to determine if they are good, and then based on that finding future visitations are allowed (p 9, section Hard focus rule). It should be noted that the *probabilities* of Chakrabarti et al. are equivalent to the claimed **statistical information.**

7.   Chakrabarti et al. teach that a focused crawler is an example-driven automatic porthole-generator. We feel that the ability to focus on a topical subgraph of the Web, as in this paper, together with the ability to browse communities within that subgraph, will lead to significantly improved Web resource discovery (p 3, last paragraph before Section 2), which meet the limitation of **wherein the statistical information-using step further comprises learning a linkage structure from at least a portion of the collected statistical information with each successive document retrieval operation such that the learned linkage structure is available for use in performing subsequent document retrieval operations requested by a user.**

8.   It should be noted that the *porthole*, which is a *subgraph of the Web*, generated by the *focused crawler* of Chakrabarti et al. is equivalent to the claimed **linkage structure** that is learned. It should further be noted that the generation of a porthole or specialized link structure (p 20, last paragraph) is equivalent to the claimed **learning a linkage structure**.

9.   Chakrabarti et al. do not explicitly teach **collecting at least a set of aggregate statistical information and a set of predicate-specific statistical information**.

10.   Chaudhuri et al. teach that the MNSA technique for determining if the existing set of statistics contains an essential set of statistics should be qualified as follows. First, note that even for a single selectivity variable, multiple statistics may be applicable with different degrees of accuracy. Second, for an SPJ query, MNSA guarantees inclusion of an essential set of the query only as long as the selectivity of predicates in the query is between g and 1-g. Third, although for SPJ queries MNSA ensures that an essential set is included among the statistics, it is necessary to extend the method beyond simple queries. Aggregation clauses can be handled by associating a selectivity variable that indicates the fraction of rows in the table with distinct values of the column(s) in the clause (Column 19, lines 35 – 63), which meet the limitation of **collecting at least a set of aggregate statistical information and a set of predicate-specific statistical information**. Because both Chakrabarti et al. and Chaudhuri et al. teach methods of collecting statistics, it would have been obvious to one skilled in the art to substitute one method for the other to achieve the predictable result of collecting aggregate and predicate-specific statistics.

11.   **Regarding dependent claim 2**, Chakrabarti et al. teach that Query construction

is not a one-time investment, because as pages on the topic are discovered, their

additional vocabulary must be folded in manually into the query for continued discovery

(p 7, lines 4 – 6), which meet the limitation of **the user-defined predicate specifies**

content associated with a document. It should be noted that the *additional*

vocabulary of pages on the topic of Chakrabati et al. is equivalent to the claimed

content associated with a document.

12.   **Regarding dependent claims 3 and 4**, Chakrabarti et al. teach that pages that

are examples associated with a topic can be preprocessed as desired by the system.

The user's interest is characterized by a subset of topics that is marked good. No good

topic is an ancestor of another good topic. Ancestors of good topics are called path

topics. Given a Web page, a measure of its relevance must be specified to the system

(p 8, lines 9 – 14), which meet the limitation of **the statistical information collection**

step uses content of the one or more retrieved documents and that **the statistical**

information collection step considers whether the user-defined predicate has

been satisfied by the one or more retrieved documents, since a determination is

made about the ancestors and preprocessed pages are used, which are equivalent to

the claimed **one or more retrieved documents**. It should be noted that the *topic* of

Chakrabarti et al. is equivalent to the claimed **content** and **predicate**.

13.   **Regarding dependent claims 5 and 6**, Chakrabarti et al. teach that we have

presented evidence in this section that focused crawling is capable of steadily collecting

relevant resources and identifying popular, high-content sites from the crawl, as well as

regions of high relevance, to guide itself. It is robust to different starting conditions, and

finds good resources that are quite far from its starting point. In comparison, standard

crawlers get quickly lost in the noise, even when starting from the same URLs (p 20,

Section 4.8 and p 18, Figure 9), which meet the limitation of **the collected statistical**

information is used to direct further document retrieval operations toward

documents which are similar to the one or more retrieved documents that also

satisfy the predicate, and that **the collected statistical information is used to direct**

further document retrieval operations toward documents which are more likely to

satisfy the predicate than would otherwise occur with respect to document

retrieval operations that are not directed using the collected statistical

information, since the focused crawling of Chakrabati et al. utilizes statistical

information (p 3) and compares their crawler to other crawlers and outlines the other's

shortcomings (Fig 9).

14.   **Regarding dependent claim 7**, Chakrabarti et al. teach that multiple citations

from a single document are likely to cite semantically related documents as well. This is

why the distiller is used to identify pages with large numbers of links to relevant pages

(p 8, last paragraph), which meet the limitation of **the collected statistical information**

is used to direct further document retrieval operations toward documents which

are linked to by other documents which also satisfy the predicate. It should be noted that the semantically related documents of Chakrabarti et al. is equivalent to the claimed **documents which are linked to by other documents which also satisfy the predicate**

15.   **Regarding dependent claim 8**, Chakrabarti et al. teach that we describe a Focused Crawler, which seeks, acquires, indexes, and maintains pages on a specific set of topics that represent a relatively narrow segment of the Web. Thus, Web content can be managed by a distributed team of focused crawlers, each specializing in one or a few topics (p 2, fourth paragraph), which meet the limitation of **the information network is the World Wide Web and a document is a web page**.

16.   **Regarding claims 10 – 17 and 19 – 26**, the claims incorporate substantially similar subject matter as claims 1 – 8, and are rejected along the same rationale.

17.   Claims 9, 18 and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Chakrabarti et al. as applied to claims 1 – 8, 10 – 17 and 19 – 26 above, and further in view of Chakrabarti et al. (Distributed Hypertext Resource Discovery Through Examples) [as cited by applicant] later referenced as Ch2 et al.

18.   **Regarding dependent claim 9**, Chakrabati et al. do not explicitly teach that **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages**.

19.   Ch2 et al. teach that other strategies are also known, such as, if the URL is of the form http://host /path, then the crawler may truncate components of path and try to fetch these URL's. If links could be traversed backward, e.g. using metadata at the server, the crawler may also fetch pages that point to the page being 'expanded.' (p 382, Column 1, lines 29 – 37), which meet the limitation of **the statistical information collection step uses one or more uniform resource locator tokens in the one or more retrieved web pages**.

20.   It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the teachings of Chakrabarti et al. with that of Ch2 et al. because such a combination would provide the users of Chakrabarti et al. with teachings of *the architecture of a hypertext resource discovery system using a relational database* (p 375, Column 1, lines 1 & 2).

21.   **Regarding claims 18 and 27**, the claims incorporate substantially similar subject matter as claim 9, and are rejected along the same rationale.

Response to Arguments

22.   Applicant's arguments with respect to claims 1 – 27 have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

23.   Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP

§ 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37

CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE

MONTHS from the mailing date of this action. In the event a first reply is filed within

TWO MONTHS of the mailing date of this final action and the advisory action is not

mailed until after the end of the THREE-MONTH shortened statutory period, then the

shortened statutory period will expire on the date the advisory action is mailed, and any

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later

than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Nathan Hillery whose telephone number is (571) 272-

4091. The examiner can normally be reached on M - F, 10:30 a.m. - 7:00 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Doug Hutton can be reached on (571) 272-4137. The fax phone number for

the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the
Patent Application Information Retrieval (PAIR) system. Status information for
published applications may be obtained from either Private PAIR or Public PAIR.
Status information for unpublished applications is available through Private PAIR only.
For more information about the PAIR system, see http://pair-direct.uspto.gov. Should
you have questions on access to the Private PAIR system, contact the Electronic
Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a
USPTO Customer Service Representative or access to the automated information
system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

NH

DOUG HUTTON
SUPERVISORY PATENT EXAMINER